Using data cloud services to manage harmful algae blooms

Aguilar, Fernando; Criado, Alberto; Monteoliva, Agustín

Harmful Algal Blooms (HABs) happen when toxic microalgae proliferate beyond control and take over rivers, lakes or ponds with costly environmental and socioeconomic impacts, for example: on fisheries, or on the availability of drinking water. At sea, this phenomenon causes red tides. Blooms and red tides are caused by a combination of meteorological, hydrodynamic and biogeochemical factors that are difficult to pin down with certainty. For these reasons, managing algal blooms is a challenge for local governments, environmental agencies and the people that depend on healthy water bodies for their livelihood. Despite the investment in waste management and monitoring systems, current methods and processes are still far from ideal. Ecohydros believes that deploying new technologies and big data analytics can pave the way for better and more efficient ways to manage harmful algal blooms. Extracting meaningful information from monitoring data is a computational challenge. The data covers hundreds of variables and parameters that need to undergo treatment, processing and analysis before they can be used in visualization tools. The predictive models also require calibration in the short and medium term in an at least semi-automatic way, performing sweeps of numerous parameters in multiple combinations, forcing the system to use high demand iterations. All this leads to an increase in the demand for computing beyond what a standard company or a standard computer center can provide.

The demo will show how cloud computing based solutions enable a system to manage not only the data ingestion from diverse sources but also the modeling of aquatic ecosystems. First of all, the data ingestion part will show how data coming from different sources are gathered (Sentinel-2, Landsat, AEMET). After downloading, data is stored in Onedata and metadata is automatically attached. These metadata are indexed and enable queries to find the datasets during the modeling part. The different datasets to be downloaded can be selected in a Jupyter notebook form (which has access to the Onedata space) as well as the type of data to find and the model to perform. When the modeling is selected, the data stored in onedata is found (using dates range and location) and used to feed the model after preprocessing. The models run using cloud computing resources and the output is also stored in onedata, so Jupyter notebook interface access directly to the generated data.